Theoretical Computer Science 739 (2018) 1–12

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Parameterized complexity and approximation issues for the colorful components problems

Riccardo Dondi[a,*], Florian Sikora[b,*]

[a] Dipartimento di Scienze umane e sociali, Università degli Studi di Bergamo, Italy
[b] Université Paris-Dauphine, PSL Research University, CNRS UMR 7243, LAMSADE, 75016 Paris, France

ARTICLE INFO

Article history:
Received 25 July 2017
Accepted 24 April 2018
Available online 2 May 2018
Communicated by V.Th. Paschos

Keywords:
Colorful components
Parameterized complexity
Algorithms
Computational biology

ABSTRACT

The quest for colorful components (connected components where each color is associated with at most one vertex) inside a vertex-colored graph has been widely considered in the last ten years. Here we consider two variants, Minimum Colorful Components (MCC) and Maximum Edges in transitive Closure (MEC), introduced in 2011 in the context of orthology gene identification in bioinformatics. The input of both MCC and MEC is a vertex-colored graph. MCC asks for the removal of a subset of edges, so that the resulting graph is partitioned in the minimum number of colorful connected components; MEC asks for the removal of a subset of edges, so that the resulting graph is partitioned in colorful connected components and the number of edges in the transitive closure of such a graph is maximized. We study the parameterized and approximation complexity of MCC and MEC, for general and restricted instances.

For MCC on trees we show that the problem is basically equivalent to Minimum Cut on Trees, thus MCC is not approximable within factor $1.36 - \varepsilon$, it is fixed-parameter tractable and it admits a poly-kernel (when the parameter is the number of colorful components). Moreover, we show that MCC, while it is polynomial time solvable on paths, it is NP-hard even for graphs with constant distance to disjoint paths number. Then we consider the parameterized complexity of MEC when parameterized by the number k of edges in the transitive closure of a solution (the graph obtained by removing edges so that it is partitioned in colorful connected components). We give a fixed-parameter algorithm for MEC parameterized by k and, when the input graph is a tree, we give a poly-kernel.

1. Introduction

The quest for colorful components inside a vertex colored graph has been a widely investigated problem in the last years, with application for example in bioinformatics [1–3]. Roughly speaking, given a vertex-colored graph, the problem asks to find the colorful components of the graph, that is, connected components that contain at most one vertex of each color. While most of the approaches have focused on the identification of a single connected colorful component. The identification of the minimum number of colorful connected components that match a given motif has been considered only in [4,5].

Here we consider a similar framework, where instead of looking for a single colorful component inside a vertex-colored graph, we ask for a partition of the graph vertices in colorful components. This approach has been proposed in bioinfor-

* Corresponding authors.
E-mail addresses: riccardo.dondi@unibg.it (R. Dondi), florian.sikora@dauphine.fr (F. Sikora).

https://doi.org/10.1016/j.tcs.2018.04.044

matics, and more specifically in comparative genomics. In this context, a fundamental task is to infer the relations between genes in different genomes and, more precisely, to infer which genes are orthologous. Genes are orthologous when they originate via a speciation event[1] from a gene of an ancestral genome. In 2011, Zheng et al. proposed a graph approach aiming to identify disjoint orthology sets, where each of such sets corresponds to a colorful component in the given graph [6] and the colorful components associated with the orthology sets are disjoint.

Different combinatorial problem formulations, based on different objective functions, have been proposed and studied in this direction [6,7]. Here, we considered two such approaches: Minimum Colorful Components (MCC) and Maximum Edges in transitive Closure (MEC). Given a vertex-colored graph, both combinatorial problems ask for the removal of some edges so that the resulting graph is partitioned in colorful components, but with different objective functions. The former aims to minimize the number of connected colorful components, while the latter aims to maximize the transitive closure of the resulting graph. A related but different problem has been considered in [2], where the objective function is the minimization of edge removal, so that the computed graph consists only of colorful components. Note that in the problems studied in this paper, the number of removed edges is never part of the objective function.

Previous results. Given a graph on n vertices, MCC is known not only to be NP-hard, but also not approximable within factor $O(n^{1/14-\varepsilon})$ unless P = NP [7]. It is easy to see that the reduction leading to this inapproximability result implies also that MCC cannot be solved in time $n^{f(k)}$ for any function f, where k is the number of colorful components. In the parameterized complexity vocabulary, it means that it is not in the XP class.

MEC is known to be APX-hard even when colored by at most three colors (while it is solvable in polynomial time for two colors), and, unless P = NP, it is not approximable within factor $O(n^{1/3-\varepsilon})$ when the number of colors is arbitrary, even when the input graph is a tree where each color appears at most twice [8]. A heuristic to solve MEC is presented in [6], while in [8], the authors present a polynomial-time $\sqrt{2 \cdot OPT}$ approximation algorithm.

Contributions and organization of the paper. In this paper we investigate deeper the complexity of MCC and MEC. More precisely, we show in Section 3 that MCC on trees is essentially equivalent to Minimum MultiCut on Trees, thus MCC is not approximable within factor $1.36 - \varepsilon$ unless P = NP for any $\varepsilon > 0$, but 2-approximable, it is fixed-parameter tractable (but not in subexponential-time) and it admits a poly-kernel (when the parameter is the number of colorful components). Moreover, in Section 4 we show that MCC is easily solvable in polynomial time on paths, while it is not in XP class when parameterized by the structural parameter Distance to Disjoint Paths.

Then we consider the parameterized complexity of MEC with respect to the number k of edges in the transitive closure of a solution. For this parameter we give in Section 5 a parameterized algorithm, by reducing the problem to an exponential kernel. We use a similar idea in Section 6, to improve it to a poly-kernel for MEC when the input graph is a tree. Finally, we show in Section 7 that results similar to those of Section 4, hold also for MEC. A preliminary version of this work appeared in [9].

2. Definitions

In this section we introduce some preliminary definitions. For any positive integer x, $[x]$ denotes the set $\{1, 2, \ldots, x-1, x\}$. Consider a set of colors $C = \{c_1, \ldots, c_q\}$. A C-colored graph $G = (V, E, C)$ is a graph where every vertex in V is associated with a color in C; the color associated with a vertex $v \in V$ is denoted by $c(v)$. If $\mathcal{C}$ is a class of graphs, the *distance to* $\mathcal{C}$ of a graph G is the minimum number of vertices to remove from G to get a graph in $\mathcal{C}$. A connected component induced by a vertex set $V' \subseteq V$ is called *a colorful component*, if it does not contain two vertices having the same color. If a graph has t connected components where each component $i \in [t]$ has exactly n_i vertices, the number of edges in its transitive closure is defined by $\sum_{i=1}^{t} \frac{n_i(n_i-1)}{2}$. In other words, for each connected component, it is the maximum number of possible edges connecting vertices of that component.

Next, we introduce the formal definitions of the optimization problems we deal with.

Minimum Colorful Components (MCC)

- **Input**: a C-colored graph $G = (V, E, C)$.
- **Output**: remove a set of edges $E' \subseteq E$ such that each connected component in $G' = (V, E \setminus E', C)$ is colorful, and the number of connected components of G' is minimized.

Maximum Edges in transitive Closure (MEC)

- **Input**: a C-colored graph $G = (V, E, C)$.
- **Output**: remove a set of edges $E' \subseteq E$ such that each connected component in $G' = (V, E \setminus E', C)$ is colorful, and the number of edges in the transitive closure of G' is maximum.

[1] A speciation is an evolutionary process from which a biological species evolves into two new species.